

METHODS AND SYSTEMS FOR AUTOMATIC DETECTION OF CONTINUOUS-TONE REGIONS IN DOCUMENT IMAGES

5 BACKGROUND OF THE INVENTION

Digital documents can be ameliorated through document processing that enhances the color, contrast and other attributes of the document. Different processing methods are used to enhance different content types. In some cases, processing that enhances one content type will degrade another content type. For
10 example, a process that enhances the legibility of text may degrade attributes of a continuous-tone (contone) image, such as a digital photograph. In mixed-content documents, different content types must be segmented in order to provide optimum document processing. Content types include contone images, halftone images, text and others. Some document types can be further delineated for more
15 optimal processing. For example, contone regions can be separated into pictorial and non-pictorial regions.

The identification and delineation of document areas by content type is an essential part of segmentation for document processing. Once each document region is segmented, each region can be processed separately according to the
20 specific needs of the content type.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1(a) is a diagram depicting the steps of some simplified embodiments of the present invention;

25 FIG. 1(b) is a diagram depicting the steps of other embodiments of the present invention;

FIG. 2(a) shows a typical mixed-content document;

FIG. 2(b) shows the local standard deviation of the luminance component of the image in FIG. 2(a);

30 FIG. 2(c) shows a mask image obtained by thresholding standard deviation with non-text regions shown in white;

FIG. 2(d) shows a mask image after morphological processing;

FIG. 3(a) shows a 64 bin luminance histogram of the input image shown in FIG. 2(a);

35 FIG. 3(b) shows the pixels found in the most populated bin of the histogram shown in Figure 3(a) for the input image shown in FIG. 2(a);

FIG. 3(c) shows the image shown in FIG. 3(b) with background value extension;

FIG. 4(a) shows a mixed content image similar to the image shown in FIG. 2(a), but printed on colored paper;

5 FIG. 4(b) shows a logic diagram for the calculation of a foreground mask;

FIG. 4(c) shows a histogram for a mixed-content page on a colored background;

FIG. 5(a) shows a lightness foreground mask where black regions correspond to background pixels and white regions correspond to foreground
10 pixels;

FIG. 5(b) shows a chroma foreground mask in a format similar to that of 5(a);

FIG. 5(c) shows a hue foreground mask in a format similar to that of 5(a);

FIG. 5(d) shows the result of an OR operation of the masks in FIGS 5 a-c;

15 FIG. 6(a) shows a final candidate region mask obtained by ANDing the local feature mask (FIG. 2(d)) and the background mask (FIG. 3(c));

FIG. 6(b) shows a final set of candidate regions after connected component labeling and bounding box computation;

FIG. 7(a) shows a luminance histogram of a spot color region;

20 FIG. 7(b) shows a luminance histogram of a contone image region; and

FIG. 8 shows elements of refinement of detected region boundaries.

DETAILED DESCRIPTION OF THE INVENTION

Referring in detail to the drawings wherein similar parts of the invention are
25 identified by like reference numerals.

Embodiments of the present invention comprise methods and systems for identification and delineation of continuous-tone (contone) regions in a digital mixed-content document. These documents comprise documents that have been digitally scanned and which contain a mixture of objects such as text, halftone,
30 pictorial contone and non-pictorial contone elements. When selective treatment of each content type is desired, content types must be segmented.

The primary motivation for segmenting pictorial contone elements is to support image enhancements that are specifically tuned to the individual image properties of the region. For example, focus, exposure, color-balance and other

attributes can be enhanced and corrected. When these content regions are segmented and enhanced separately, each region of the document can be enhanced without deleterious effects on other regions.

To properly detect pictorial content regions, two key discriminations should be made: 1) distinguishing text from content regions and 2) distinguishing pictorial from non-pictorial content regions. The issue with text is that it shares many local properties with content regions - e.g., large local variance at edges and local uniformity in interior areas. Since content image enhancements are highly detrimental to text legibility, it is beneficial to eliminate all false-positive classifications of text as content. The issue with non-pictorial content regions has mainly to do with the cost of the proposed image enhancements and the fact that such computation, if not deleterious, would be largely wasted on non-pictorial regions. Therefore, eliminating non-pictorial false-positives is highly desirable as well.

Embodiments of the present invention comprise methods and systems for segmentation of image content types. These types comprise text, background, pictorial content, non-pictorial content and others. Image element attributes such as luminance, chrominance, hue and other attributes may be used to detect and delineate image content types. In some embodiments of the present invention, only the luminance component of the input image is utilized. In other embodiments the chrominance components and/or color and hue attributes, as well as other attributes, may be used to improve the accuracy or other performance characteristics of the algorithm.

Embodiments of the present invention may process many types of digital images. These images may be obtained from a scanner, from a digital camera or from another apparatus that converts physical media into a digital file. These images may also be originally generated on a computer such as through the use of a graphics application as well as by other methods. A typical mixed-content image is shown in Figure 2(a).

Some embodiments of the present invention may be described with reference to Figure 1(a). In these embodiments, image element attributes are obtained by known methods, such as by reading an image file. These attributes may comprise pixel luminance and chrominance values as well as other information. Image element attribute data may be filtered to remove noise and/or

downsampled to make the image data easier to analyze and store.

In these embodiments, image data is analyzed to determine whether text regions are present in the image 42. This may be performed by local computation of a discriminating feature such as standard deviation, spread or some other feature.

- 5 Once text regions are located, they are bounded and tagged as text or non-contone regions.

Background regions are also detected 44 and tagged by methods explained below as well as other methods. These methods may comprise analysis of luminance histogram data including a determination of the histogram bin

- 10 containing the maximum number of image pixels when this value is above a threshold value. An analysis of neighboring histogram bins may also be used to modify the background detection routine.

Once text and background regions are found, these regions may be combined 46 and eliminated from consideration in the contone detection process.

- 15 If no background or text is found, the entire image may be tagged as a contone region 48.

When background and text regions are found, the remainder of the image may be analyzed to identify contone regions 50. This analysis may comprise an analysis of the region's luminance histogram data. As contone regions typically have a

20 uniformly-distributed histogram, this feature may be used to identify these regions. In some embodiments, the number of populated histogram bins whose pixel count exceeds a threshold value is compared to a bin number threshold value. When the number of bins exceeds this value, the regions is considered a contone region.

- In some embodiments, this determination is subject to modification in a
- 25 secondary determination using regional properties. In these embodiments, regional properties, such as region area and luminance distribution are considered. If a region's area is smaller than a particular area threshold value, the region may be removed from consideration as a contone region. In some
- embodiments, the area threshold value may be related to a page characteristic,
- 30 such as page width. In particular embodiments, the area threshold value may be equal to the square of one tenth of the page width.

A further regional property may be used to further identify a contone region as a pictorial contone region. In these embodiments, the luminance histogram data is analyzed. The two dominant histogram bins are removed from

consideration and the remaining bins are analyzed to determine whether they represent a typical bi-modal distribution. If the distribution is more varied than bi-modal, the region is considered pictorial.

Once initial regions have been identified, embodiments of the present invention may recursively analyze these regions to identify sub-regions that may exist within each region. This process may continue recursively until each sub-region is homogenous.

Some embodiments of the present invention may be described with reference to Figure 1(b). In these embodiments, the input image may be optionally downsampled to generate a low-DPI (dots-per-inch) version, in order to significantly reduce the memory overhead of the algorithm and to reduce the required calculation resources. Some embodiments may not downsample when sufficient processing resources are available. After downsampling or the omission thereof, the input image is processed to reduce the effects of noise.

In some embodiments, a 3x3 median filter is used to process the image; however, alternative filtering or image processing methods can also be used to reduce the noise in the image data. After the pre-processing step, a discriminating feature is computed locally for each pixel to highlight and subsequently identify the text regions in the image. In some embodiments, the discriminating feature is the standard deviation, calculated for each pixel using a 5x5 window.

Other embodiments can utilize alternative local features, such as the spread, which is defined as the number of pixels in a local window not equal to the maximum or the minimum values in the window.

Figure 2(b) shows the local standard deviation feature computed for the sample document image in Figure 2(a). It is clear in Figure 2(b) that the standard deviation values tend to be higher in the text regions of the document image.

This property is exploited through a thresholding operation to discard the text areas in the image and locate a set of candidate regions on the page that may correspond to continuous-tone content. Figure 2(c) shows the mask generated from the standard deviation image by thresholding; the white areas in the mask correspond to non-text locations whose standard deviation value is below the predetermined threshold T_T . The value of T_T can be chosen in different ways. In some embodiments, T_T is set to 32 for all input content. After thresholding, the mask image is processed using morphological methods such as erosion and

opening to eliminate small, isolated regions. The final mask image obtained after morphological processing is shown in Figure 2(d).

The initial mask of candidate continuous-tone regions can be further improved by identification and removal of the background pixels in the document image. In some embodiments of the present invention, the range of gray level values that correspond to the background can be determined through analysis of the luminance histogram of the document region. The main assumption is that the background pixels comprise a considerable portion of the region of interest, which in turn implies that these pixels constitute a significant peak in the luminance histogram.

To detect the document background, the gray level values in the luminance histogram that correspond to the bin with the maximum number of pixels are taken as the initial estimate of the region background. The pixel count in the selected bin must exceed a predetermined threshold to be classified as a potential background; otherwise, it is determined that no distinct background exists for the region of interest. If no distinct background exists, the entire region may be labeled as contour. The background detection threshold T_B can be set in various ways. In some embodiments, T_B is computed as 12.5% of the input image size. After the initial range estimate for the background is computed, it may be further expanded through analysis of the neighboring bins in the histogram. This latter stage renders the algorithm more robust to the effects of noise, blur due to scanner characteristics, and so on. Furthermore, the background detection method is able to identify background regions of any color, since no assumption is made during histogram analysis on where the largest peak may be located. The number of bins N_B used to construct the luminance histogram may vary. In some current embodiments of the invention, N_B is set to 64. The background mask image is finally processed using morphological methods such as erosion and opening to eliminate small, isolated regions.

Figure 3 illustrates the background detection process of some embodiments in more detail. The document image of interest is shown in Figure 2(a). Figure 3(a) shows the luminance histogram H_{lum} computed for the entire page. As seen in Figure 3(a), the bin with the maximum number of pixels in H_{lum} is bin no. 64, which corresponds to the gray level range [252,255]. The document pixels that correspond to this range of values are depicted in Figure 3(b) in black. The range

is then progressively expanded at either end, by incrementing (decrementing) the upper (lower) bound of the range by 1, and determining if the number of pixels added through this operation is sufficient. If the number of pixels added through the expansion exceeds a predetermined threshold, the background range is
5 increased to include the new gray level value; otherwise, the expansion process is terminated. Figure 3(c) shows the document pixels that correspond to the final background range.

Methods that use only luminance channel information about a document page may be throwing away valuable information that could be used to further
10 narrow the background pixels specification. To try and take advantage of the information carried by the color channels, another set of embodiments has been developed that may use luminance, chroma and hue channels. These alternative embodiments are illustrated in Figure 4 with results of the various steps shown in Figure 5. The method first calculates a luminance channel histogram similar to the
15 previous embodiment shown in Figure 2. The peak of this histogram is assumed to correspond to the main background luminance level. A region is then carved out from the histogram based on noise and MTF measurements to identify the luminance levels that are likely to still belong to the background. Figure 4(c) illustrates the luminance histogram for the document shown in Figure 4(a) and the
20 area between the vertical lines 56 & 58 corresponds to the luminance levels of the background. Only the pixels denoted as background in Figure 5(a) are run through a similar process for chroma and hue channels. This narrows the definition of the background and allows more foreground pixels mislabeled by the luminance criterion to be re-added to the foreground in the final mask denoted in
25 Figure 5(d).

The binary masks obtained through background detection and local analysis are then merged 20 through an AND operation to yield the set of candidate continuous-tone regions. The merging process allows us to eliminate the text regions completely, to reduce the number of candidate locations and to
30 refine their boundaries. Figure 6(a) shows the results of the AND operation on the masks. Connected component labeling 24 is then performed on the mask to determine the number and locations of the connected image regions. Finally, the bounding box of each labeled region is computed to generate the final mask, as depicted in Figure 6(b). The shaded regions in the mask correspond to the non-

text, non-background regions that will be further analyzed and labeled as 'contone' or 'non-contone'.

Once the candidate regions in the document are identified, the luminance histogram of each region is inspected to decide whether the region is continuous-tone 26. The decision is typically made based on regional histogram uniformity. A continuous-tone image region is expected to have a fairly uniform gray level distribution, whereas a spot color area often has a heavily skewed histogram. The uniformity of the region histogram is established by counting the number of "populated" bins N_{pop} (i.e., those bins whose pixel count exceeds a threshold T_P) and comparing this total to a predetermined threshold T_C . If N_{pop} exceeds T_C , the region is classified as a continuous-tone region; otherwise, it is labeled as non-contone.

In some embodiments, the candidate regions denoted in Figure 6(b) are verified using region properties. In some of these embodiments, these properties are area and luminance distribution. Small regions labeled by the local measures as contone are more likely to be mislabeled than are large connected regions. Region area may be used to either identify or remove text elements. These areas may be removed, in some embodiments, using a threshold T_A derived from the page width as $(0.1 \times \text{page width})^2$. The second property is used to find pictorial contone regions.

Figure 7 shows the luminance histograms for a spot color region (a) and pictorial region (b). The histogram for the spot color shows the values are largely located around a single peak, while the histogram of the pictorial region is more evenly distributed. Some embodiments of this verification procedure eliminate the histogram counts around the two largest peaks in the histogram. The pictorial contone verification procedure eliminates the counts of the pixels around the two dominant peaks for the region histogram then sums the remaining bins to determine if a significant number of pixels don't belong to a bi-modal distribution. In some embodiments, the bi-modal distribution model used is one from studying the luminance distribution from inverted text samples. If the region has enough luminance levels outside of this model, then the region is considered pictorial contone. In Figure 7(b) the regions between the lines centered around the peaks correspond to the bins eliminated from the summation.

After classification 26, the boundaries of the identified continuous-tone

regions need to be refined and more accurately located. Due to the initial downsampling operation and subsequent morphological processing, the bounding boxes of the detected regions do not correspond to the correct region boundaries in the original image. To refine the boundaries, the detected bounding box
5 coordinates are first projected to the original image size 28. The size of the bounding boxes may be reduced by a fixed amount, to ensure that the detected bounding box is always smaller than the actual bounding box of the region. Each side of the detected bounding box is then expanded outward, until a termination criterion is met. In some embodiments, the termination criterion compares the
10 pixels added in each step of the expansion process to the detected background values. If the number of background pixels in the added row/column exceeds a certain number, expansion in that particular direction is terminated. Figure 8 depicts how bounding box refinement is accomplished.

Once bounding box refinement is completed, the detected contone regions
15 are recursively processed using the steps described above, until no valid background is identified for a given region. The continuous-tone detection processes are applied to (sub)regions in the document image recursively until a predetermined termination criterion is met. The recursive approach enables the algorithm to handle multiple and nested local background regions in the document,
20 thereby allowing the accurate detection of the boundaries of all continuous-tone content on the page.

The detailed description, above, sets forth numerous specific details to provide a thorough understanding of the present invention. However, those skilled in the art will appreciate that the present invention may be practiced without these
25 specific details. In other instances, well known methods, procedures, components, and circuitry have not been described in detail to avoid obscuring the present invention.

All the references cited herein are incorporated by reference.

The terms and expressions that have been employed in the foregoing
30 specification are used as terms of description and not of limitation, and there is no intention, in the use of such terms and expressions, of excluding equivalents of the features shown and described or portions thereof, it being recognized that the scope of the invention is defined and limited only by the claims that follow.